DOCUMENT RESUME

ED 328 925 CS 507 381

AUTHOR Dudczak, Craig A.; Day, Donald
TITLE The Impact of Paradigm Consistency on Taxonomic Boundaries in CEDA Debate.
PUB DATE Nov 90
NOTE 41p.; Paper presented at the Annual Meeting of the Speech Communication Association (76th, Chicago, IL, November 1-4, 1990).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Communication Research; *Debate; *Evaluation Criteria; *Judges; *Models; Questionnaires; Speech Communication; Surveys
IDENTIFIERS Cross Examination Debate Association; *Judge Philosophy Statements; Paradigmatic Responses; Paradigm Shifts

ABSTRACT

A study reported on two experiments which addressed the question of whether debate judges do as they say they will with regard to the advent of judge philosophy statements. The larger goal of the combined experiments was to discover whether: (1) judging paradigms operate meaningfully in Cross Examination Debate Association (CEDA) debate and (2) what elements these paradigms contain. The first experiment analyzed the correspondence among critic preferences expressed through 23 judge philosophy statements, responses to a survey instrument, and comments/decision criteria expressed on debate ballots. The second experiment analyzed the consistency between 39 critics' responses to a questionnaire and their evaluations on the template portion of ballots. Three research questions and nine hypotheses were studied in these two experiments. Results showed little reliability for the questionnaire as a predictor of critics' ballot behavior. Paradigm preferences showed limited association between professed paradigms and subsequent ballot behavior. Results also indicated that traditional paradigms largely overlap each other, reducing paradigm distinctiveness. The nine hypotheses showed limited, insignificant differences between critics grouped by metaparadigm categories. (One figure and five tables of data are included. Appendixes include: Syracuse debate union judging criteria questionnaire, coding categories for ballot comments, and judge philosophy coding categories. Seventeen references are attached.) (MG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED328925

The Impact of Paradigm Consistency on Taxonomic Boundaries in CEDA Debate (Revised Manuscript)*

Craig A. Dudczak
Speech Communication
Syracuse University

Donald Day
Information Studies
Syracuse University

Speech Communication Association Conference
American Forensic Association
Chicago, Illinois
November 1990

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Craig Dudczak
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS507381

* Revisions to text were minor editing. (January 1991)

Running Head: Taxonomic Boundaries in CEDA

The Impact of Paradigm Consistency on
Taxonomic Boundaries In Ceda Debate

The advent of judge philosophy statements in academic debate is predicated upon the assumption that debate critics would formulate their decision criteria better by articulating them beforehand. This also would afford debaters an opportunity to adapt to their critics' expressed preferences. While a number of studies have evaluated critics' paradigm preferences in NDT (Cox 1974; Cross & Matlon 1978; Thomas 1977) and in CEDA (Buckley 1983; Lee, Lee & Seeger 1983), these surveys have not established whether expressed preferences actually are used in judging debates. Judging philosophies and survey responses may be taken as "ought" statements; statements by critics of how they believe they "would" evaluate a debate. However, unless confirmed by decision criteria actually employed in debate rounds, philosophies may fail to represent meaningful differences in judges' preferences to which debaters can adapt. Without such confirmation, the utility of judge philosophy statements in academic debate is open to question.

The present study reports two experiments which address the question of whether judges "do as they say they will." The larger goal of the combined experiments is to discover whether (1) judging paradigms operate meaningfully in CEDA debate and (2) what elements these paradigms contain. The first experiment analyzes the correspondence among critic preferences expressed through judge philosophy statements, responses to a survey instrument, and comments/decision criteria expressed on debate

ballots. The second experiment analyzes the consistency between critics' responses to a questionnaire and their evaluations on the template (top) portion of ballots.

This investigation is justified by the scarcity of research regarding debate critic decision criteria. Early investigations (Cox 1974; Cross & Matlon 1978; Thomas 1977; Buckley 1983; Lee, Lee & Seeger 1983) surveyed critic paradigm preferences through self-report instruments. These surveys were limited to indicating "professed" beliefs, since they were not intended to validate the extent to which preferences actually were applied. More recent work by Gaske, Kugler and Theobald (1985) attempted to discriminate among CEDA judging paradigms, but relied upon unequal (and generally subcritical) cell sizes (61-65). Brey (1989) analyzed CEDA philosophy statements to discover the elements of judge preference, but his analysis did not indicate whether paradigm preferences correlated with discernible patterns of judging behavior.1

Even less research has focused upon the artifacts of debate evaluation. Bryant (1983) conducted a content analysis of NDT and CEDA debate transcripts to compare evidence use within each format.2 Hollihan, Riley, and Austin (1983) used content analysis of NDT and CEDA ballots to determine thematic "visions" embraced respectively within these two debate formats. While their analysis of ballots suggested that different visions are held by NDT critics versus CEDA critics, without knowledge of the critics' prior attitudes (as demonstrated through judging philosophies, for example), one cannot know whether ballot

comments reflected critic preference or circumstances unique to debate rounds.3

There were only three research reports that compared judge philosophy statements with ballot artifacts. Henderson and Boman (1983) reported high consistency (83.5%) between a set of NDT judge philosophy statements and corresponding ballot comments, although their analytic procedures make their findings suspect.4 Dudczak and Day (1989a) found lower consistency (54.9%) in a pilot study of CEDA critics.5 They reported that critics' claims that they felt "evidence out of context" and "quality of analysis" correlated about 70% of the time with the actual likelihood that these critics would apply "evidence of context" as a voting issue. Dudczak and Day also reported that several clusters of paradigms were correlated with decision criteria cited in critics' ballots.6

A secondary analysis of Dudczak and Day's pilot data (1989b) sought to isolate differences among traditional paradigms. Paradigm boundaries were found to be porous and unreliable. The willingness of 94 percent of critics to apply a paradigm other than their professed preference (if asked to do so by debaters) diminish the usefulness of paradigm preference statements. Support for distinctions among paradigms was found only for Argument Critic and Stock Issues paradigms. Even in these instances, support was relatively weak.

Taken as a whole, the literature on judging paradigms is limited to the mere existence of preferences, with weak and

inconsistent evidence connecting preferences to actual use. Since ballots constitute the primary feedback for debaters, an attempt should be made to describe decision criteria in a more systematic fashion employing actual artifacts (i.e., ballots).

The study in progress extends the analysis reported in the pilot study (Dudczak and Day 1989a; 1989b). A Number of experiments were designed to assess the relationships among judge philosophy preferences, critic preferences (as measured through a survey questionnaire), and critic behavior (as measured through judges ballots). Two experiments are reported in this manuscript.

EXPERIMENT #1

Three research questions were evaluated and four hypotheses tested in this experiment.

Q1: What is the strength of the relationship between professed reasons for decision as claimed in a questionnaire and actual reasons for decision cited in debate ballots?

The pilot study (Dudczak and Day 1989a) revealed two instances in which professed preferences from a questionnaire correlated with reasons for decision cited on ballots. "Evidence out of context" cited as important in survey responses correlated reasonably well ($r = .699$) with its mention on ballots. Critics' survey preferences for "quality of analysis" correlated similarly with ballot comments regarding "evidence out of context" ($r = .698$). The present study would be expected to confirm these results and to determine whether other preferences were strongly associated with ballot comments.7

Q2: What is the strength of relationship between professed judging paradigms as claimed in a questionnaire and reasons for decision cited in debate ballots?

Pilot study results (Dudczak and Day 1989a) indicated that several clusters of ballot behavior were characteristic of specific distinct paradigms. Critics who claimed Tabula Rasa, Value Comparison, Argument Skills, Hypothesis Tester, Judicial Model, and Argument Critic paradigm preferences were about equally likely (range = .698 to .685) to cite "evidence out of context" in decisions. Similarly, Value Comparison, Argument Skills, Judicial Model, and Argument Critic judges were relatively consistent (range = .674 to .644) in their application of "counterintuitive arguments" in decisions. Finally, Judicial Model and Argument Critic judges were similar (range = .589 to .553) in citing "quality of analysis" as a discriminant. The current study expected to confirm these results (and to identify other paradigm clusters).

Q3: Which traditionally recognized paradigms are sufficiently distinct in terms of decision criteria to stand alone as taxonomic elements and which should be merged with others based upon actual ballot behaviors?

Analysis by Dudczak and Day (1989b) indicated that four pairs of traditional paradigms were sufficiently similar to be considered potential combined profile types (Value Comparison - Argument Critic; Argument Skills - Argument Critic; Argument Critic - Hypothesis Tester; and Stock Issues - Judicial Model). Argument Critic and Stock Issues paradigms were the only

traditional paradigms that displayed sufficient distinctiveness to be considered unique.

Only limited sets of characteristics were identified with any of the paradigms. None of the candidate profile types correlated more strongly with key discriminators than did traditional paradigms: only the Stock Issues paradigm showed a (moderately strong) correlation to key discriminants.

The three research questions were intended to identify characteristics of avowed critic preference as measured through consistencies among paradigm types, philosophy statements, survey responses and ballot comments. Four hypotheses previously tested by Dudczak and Day (1989a) also were replicated in the current analysis.8

H1: The mean proportion of presentational (vs. substantive) remarks on ballots by Audience-centered critics (Argument Skills, Argument Critic, Public Audience) will be greater than the proportion of such remarks made by Analytic-centered (Value-Comparison, Policy Implications, Stock Issues, Hypothesis Testing, and Judicial Model) critics.

H2: The mean proportion of ballots devoted to critique (vs. decision criteria) by Audience-centered critics will be greater then the proportion allotted by Analytic-centered critics.

H3: The mean proportion of ballots devoted to decision criteria (vs. critique) on elimination round ballots will be greater than the proportion allotted in preliminary rounds.

H4: The mean proportion of substantive (vs. presentational) remarks made on elimination round ballots will be greater than the proportion of such remarks made in preliminary rounds.

Pilot study results failed to prove the first two hypotheses, although the data were in the anticipated direction.

Hypotheses #3 and #4 both were found to be significant in pilot results ($p = <.05$).9 We expected to find that the national sample used in the current analysis would support the first two hypotheses more strongly than did the (regional) pilot sample, and would reconfirm the remaining hypotheses.

Method

The current study integrated structured data (from the questionnaire and template [top] portions of ballots) with unstructured data (from judging philosophies and ballot comments). The use of survey research in concert with content analysis can yield complementary findings which are more valid than those obtained using either alone (Paisley 1969; Webb and Roberts 1969). Structured data limit respondents' choices to those dictated by the researcher. Content analysis, on the other hand, begins with a view of reality held by the subject and attempts to conform that perspective to the analytic scheme of the researcher (Holsti 1969; Krippendorff 1980).

Subjects:

Subjects used in the study were debate critics who judged debate rounds at CEDA tournaments during the Fall 1989 season. Most subjects had previous experience as debaters (90.9%) although almost half (43.8%) had two or fewer years' judging experience. For a subject's work products and instrument to be included in this part of the study, s/he must have completed a judge philosophy statement and survey questionnaire, plus a minimum of six ballots written for the Fall 1989 CEDA topic.

Eighty-seven subjects completed the questionnaire. Philosophy statements for forty-two of these respondents were gathered from the CEDA Judge Philosophy Handbooks or solicited at several tournaments.10 Ballots in sufficient numbers for analysis (six or more per critic) were available for one hundred and eighteen critics (only twenty-three of whom had completed both a philosophy statement and a questionnaire). Hence, twenty-three sets of subject responses were used for analysis in this experiment.

Materials:

The work products and instrument examined in this study included 1) judging philosophies, 2) ballots completed during competition at tournaments, and 3) a structured questionnaire administered at tournaments (following a majority of the rounds). Each of the three measures had an unique development history.

Questions for the survey were drawn initially from the researchers' personal experiences at various levels of debate. The initial pilot study (Dudczak and Day 1989a; 1989b) revealed a need for additional criteria for decision and for inclusion of valences for all decision elements. Two questions were taken from Buckley (1983). The sequence of questions and style of respondent selection options were based upon professional marketing experience and coursework in survey research techniques.

The coding of worksheets for content analysis of philosophy statements and ballots included the use of matrices to capture

the proportion of presentational vs. substantive elements noted and the degree of critique vs. decision criteria appearing in critics' written comments. Coding forms used for the pilot study were expanded to include new discriminants and a coding category description form was drafted to standardize discriminant boundaries for coders. Worksheets adopted the list of traditional paradigms employed by Buckley (1983).

The one instrument and two work products used in the study may be visualized in a two-by-two table. Both the philosophy and questionnaire are normative--"ought"--documents; the ballots are applied documents. The philosophy and comment portions of ballots are unstructured; the questionnaire and template (top) portions of ballots are structured. Using these distinctions, future studies may examine content, construct, and predictive validity of these types of documents.

FIGURE 1

Construct and technique matrix of tools in the study

Procedure:

A two-page questionnaire was used incorporating 32 Likert Scale items, five yes/no selections, five multiple option questions, two single selection choices, one 10-item rank order question, and two 3-item proportional weighting scales. The questionnaires were administered to judges at CEDA debate tournaments. Twenty-eight of the Likert Scale items also asked whether the operation of an element in a round would help or hurt the team involved.11

Twenty-nine tournament directors who had hosted CEDA tournaments during the Fall 1989 season were asked to administer the questionnaire. Sixty-nine questionnaires were returned from eleven tournaments; two additional questionnaires were returned directly by respondents. A follow-up solicitation mailed to critics yielded an additional sixteen questionnaires. A total of eighty-six completed surveys were obtained.12

Official ballots submitted by judges at eleven (of the twenty-nine) CEDA tournaments comprised the second source of data. Each round was considered an unique case for purposes of statistical analysis. Of the 1653 ballots returned, 1519 were usable.13 Only the usable ballots for the twenty-three subjects who had a minimum of six ballots each (and who had completed a philosophy statement and a survey) were included in this portion of the study (N = 217). Ballot comments were recorded on a standardized coding form.14

The third source of data was judge philosophy statements,

already described. Judge philosophy statements were rated independently by two coders. Of the forty-two items on the judge philosophy coding form, ten were binary, thirty were three category choices, and two were ten-category choices. The overall inter-coder reliability was (r = .492), although the method of calculating reliability avoided conventions that would have inflated reliability.15 Table 1 reports the discriminants for which relatively high reliability levels warrant further investigation.

Table 1

Discriminants Revealing High Inter-coder Reliability:
Judge Philosophy Statements

DISCRIMINANT	INTER-CODER RELIABILITY
Tabula Rasa	1.000
Judicial Model	.691
Hypothesis Testing	.585
Uniqueness	.935
Obnoxious Behavior	.894
Counterwarrants	.738
Burden of Rejoinder	.683
Ethics	.585
Substantive Issues	.563

Data processing for the study was performed on an IBM PC using PC-FILE PLUS (a database program) and on an IBM 3090 Mainframe using SAS (a statistical package). Data were entered via PC-FILE, converted to standard data format (SDF), manipulated using BASIC programs written for this study, then uploaded to the mainframe to SAS univariate and correlation runs.

Results

Research question 1 asked the strength of relationship between reasons for decision professed on a questionnaire and the actual reasons for decision cited in debate ballots. Univariate analysis revealed identified seven ballot discriminants that appeared to be associated with questionnaire discriminants. However, there appeared to be little association between respondents' rating of items in the questionnaire and their subsequent ballot comments (Table 2). No correlation approached the levels observed in the pilot study.

Table 2

Correlation Between Questionnaire Items and Ballot Comments

QUESTIONNAIRE DISCRIMINANT	BALLOT DISCRIMINANT						
	Topic	Justif	Organ	Criter	EvSuf	CrossX	DropAr
Signif	.042	-.271	-.137	-.081	.160	-.160	-.269
PresentSkl	-.282	.103	-.147	-.069	.054	.119	-.089
EvidAttack	-.021	-.149	-.077	-.166	.175	.219	-.142
EvidContxt	-.018	.163	-.112	-.107	-.117	.095	.219
EvidSuffnt	-.136	-.001	-.282	.051	.045	-.008	.227
EvidApply	-.121	-.126	-.183	-.087	-.008	-.029	-.017
Topicality	.111	-.114	.126	-.099	.044	.168	-.174
QualAnalys	-.047	.155	-.171	-.081	-.159	.142	.205
NoValue	-.052	-.019	-.246	-.013	-.159	-.205	-.015
TheoryArg	.022	-.081	.146	-.062	-.049	-.061	.188
DroppedArg	.014	-.184	-.052	.129	.069	-.236	.058
Justifica	-.218	-.298	-.074	-.080	.227	.102	-.121

Research question 2 asked whether critics' professed judging paradigms had more than a chance relationship with the reasons for decision cited in their ballots. The only correlation which

merits further investigation was an association between the Stock Issues paradigm and the appearance of "justification" on ballots (r = .347). Table 3 presents the correlation matrix between the seven ballot discriminants and the nine paradigms that respondents were asked to rank on the questionnaire.16

Table 3

Correlation between Judge Paradigms and Ballot Comments

PARADIGM	BALLOT DISCRIMINANT						
	Topic	Justif	Organ	Criter	EvSuf	CrossX	DropAr
ArgCrit	.022	-.239	.225	.063	-.054	.081	-.247
ArgSkil	.153	.175	-.140	.131	-.110	-.201	.238
PubAud	.179	.051	.195	.091	.011	-.151	.164
HypoTst	-.049	-.108	.037	-.024	.119	-.229	.019
Tabrasa	.102	.140	.130	-.129	.052	-.015	.049
Valcomp	.196	.119	-.067	.205	-.124	.015	.018
Judical	.069	-.039	.103	.003	-.152	-.179	.010
PoluImp	-.145	.166	-.233	-.010	-.098	.079	-.016
StokIsu	-.176	.347	-.261	.119	-.249	-.025	.089

Research question 3 asked whether traditionally recognized paradigms are sufficiently distinct or whether elements of some paradigms should be merged to create new paradigms, based on critics' ballot behavior. The nine traditional paradigms were matched against the seven ballot discriminators to reveal potential patterns of similarity and difference. The pairing of paradigms on shared characteristics (for the seven key discriminators) revealed a pattern of commonality. Table 4 reports the matched pairs.

Table 4

Commonality of Correlations Among Paradigms on Key Discriminators

NUMBER OF MATCHES PER PARADIGM PAIR

	TR	VC	PI	AS	AC	SI	PA	HT	JM
TR	--	7	5	5	6	4	5	6	5
VC		--	5	4	3	4	4	5	5
PI			--	3	6	4	4	4	4
AS				--	5	3	6	5	4
AC					--	4	5	4	3
SI						--	4	4	4
PA							--	3	4
HT								--	6
JM									--

Note1: TR = Tabula Rasa; VC = Value Comparison; PI = Policy Implications; AS = Argument Skills; AC = Argument Critic; SI = Stock Issues; PA = Public Audience; HT = Hypothesis Testing; JM = Judicial Model

Note2: Pairs considered atypically similar in terms of key discriminators had a difference of no more than 0.1 correlation on at least six of the seven discriminators

The low differences among correlations obtained for key discriminators indicated minimal paradigm distinctiveness. Nevertheless, when six of seven or more of the discriminators fail to distinguish greatly among paradigms, there is evidence to suggest that a merger of traditional paradigms had occurred. The following candidate paradigm pairs had six or more atypical similarities on the seven key discriminators:17

Tabula Rasa - Value Comparison
Tabula Rasa - Argument Critic
Tabula Rasa - Hypothesis Tester
Policy Implication - Argument Critic
Argument Skills - Public Audience
Hypothesis Tester - Judicial Model

These pairs are candidates for further research. This phenomenon suggests that traditional paradigms may not be distinctive enough to delineate unique judging behaviors.

Hypothesis #1 proposed that the proportion of presentational (vs. substantive) remarks on ballots by Audience-centered critics would be greater than that for analytic-centered critics. No significant correlation was found between the characterization of a critic as audience-centered and the likelihood of presentationally oriented remarks appearing on his or her ballots ($r = .072$). The characterization of a critic as audience-centered showed a slightly stronger correlation with substantive remarks on ballots. A similarly weak relationship between analytic-centered critics and presentational comments was obtained ($r = -.042$). The strongest association found was a .31 correlation (in the expected direction) between analytic-centered critics and the incidence of substantive comments on ballots.

Hypothesis #2 proposed that audience-centered critics would devote more of their ballots to critique rather than decision criteria compared to analytic-centered critics. The results were in the predicted direction, but failed to attain significance. Analytic-centered critics were more inclined to devote the greater proportion of their ballots to decision criteria; they were nearly equally disinclined to include critiques. Table 5 summarizes the association between meta-paradigm types and the proportion of ballots taken by comments.

Table 5

Correlation between Critic Type and Comments

META-PARADIGM	COMMENT TYPE	
	Critique	Decision
Audience-centered	.101	-.134
Analytic-centered	-.223	.249

Note #1 ($F' = 1.17$ w/ 48 and 129 DF, $p > F' = 0.4913$)

Hypothesis #3 proposed that the mean proportion of ballots devoted to decision criteria (vs. critique) would be greater in elimination rounds than in preliminary rounds. Results showed no correlation of any merit to support this prediction. The maximum value ($r = .079$) obtained suggests little difference between critics' preliminary and elimination round ballots.18

Hypothesis #4 predicted that the mean proportion of substantive (vs. presentational) remarks made on elimination round ballots would be greater than this proportion for preliminary rounds. Results showed very little support for the hypothesis, except a minor indication that elimination rounds do feature fewer presentational elements ($r = 0.140$). This relationship is in the predicted direction, but with very weak support.19

EXPERIMENT #2

The focus of experiment 2 was to compare critics' professed preferences with the evaluations on template portions of ballots.

The ballot template requires structured responses, unlike the written section of ballots (for which the critic has complete latitude to write any comments or decision criteria). Five hypotheses were tested in this experiment:

H5: Analytic-centered critics award more speaker points than do audience-centered critics in preliminary rounds.

The assumption operating here was that audience-centered critics view "speaker" points more literally than do analytic-centered critics,20 who view speaker points as "global" evaluations of debaters' performance in the round (Hollihan, Riley, and Austin 1983). Pilot results for hypotheses #1 and #2 were consistent with this hypothesis, although they did not attain significance.

H6: Analytic-centered critics record a greater proportion of low-point wins than do audience-centered critics.

H7: Critics with relatively more NDT experience are more likely to record low-point wins.

Each of the preceding hypotheses assumed different "visions" between Analytic- and Audience-centered critics. NDT-experienced critics have been acculturated to different functions for debate. Most broadly stated, analytic-centered critics were expected to discount presentational skills. In the circumstance where a single key issue is defaulted, they should find it easier to resolve a decision exclusively on an analytic ground.

H8: The difference in speaker points between winning and losing teams is less for analytic-centered critics than for audience-centered critics.

H9: The difference in ranks between winning and losing teams is less in rounds judged by analytic-centered critics than in those judged by audience-centered critics.

The authors' anecdotal experience suggests that analytic-centered judges tend to see rounds as closer, therefore feel that debaters deserve nearly equal points and ranks.

Method

Structured data from the template portions of ballots were compared to structured data from the questionnaire. Questionnaires provided information about critics' perceived preferences, preferences that presumably were germane when they had completed the top portions of ballots. Critics' expressed preferences were compared to actual ballot behavior.

Subjects:

Subjects were debate critics who judged at Fall 1989 CEDA tournaments. Eighty-seven subjects completed a questionnaire on judging preferences. Thirty-nine of the judges who completed the questionnaire also wrote six or more ballots. These thirty-nine judges constituted the subjects for this experiment.

Materials and Procedures:

The questionnaire and procedures were described previously. Subjects already had completed the questionnaire; the template portions of ballots were coded and recorded.

Results

Hypothesis #5 proposed that analytic-centered critics would award more speaker points than would audience-centered critics.

Results showed no significant difference between these two categories of critics in terms of the number of points they typically award.

Hypothesis #6 predicted that analytic-centered critics would be more inclined to award "low-point" wins than would audience-centered critics. As an ancillary prediction, hypothesis #7 proposed that critics with previous NDT experience would be more likely to award low-point wins. Neither of these hypotheses was supported. Analytic-centered critics were somewhat more inclined ($r = .126$) than audience-centered critics ($r = -.053$) to award low-point wins, though the result was not significant. While previous NDT experience was associated modestly with low-point wins ($r = .101$), it also was not significant.21

Hypotheses #8 and #9 (respectively) predicted that analytic-centered critics would award lower range differences in (1) speaker points and (2) speaker ranks between winning and losing teams than would audience-centered critics. None of these predictions were supported. The only finding observed in the predicted direction was that analytic-centered critics were associated somewhat with less difference in speaker ranks between winning and losing teams ($r = -.121$). However, this finding was not significant.

DISCUSSION

Three research questions and nine hypotheses were studied in two experiments. Results showed little reliability for the questionnaire as a predictor of critics' ballot behavior.

Paradigm preferences in research question #2 showed limited association between professed paradigms and subsequent ballot behavior. Research question #3 indicated that traditional paradigms largely overlap each other, reducing paradigm distinctiveness. The nine hypotheses showed limited, insignificant differences between critics grouped by meta-paradigm categories.

The two experiments showed less significant results than similar studies in the two preceding pilot studies. The balance of this discussion section explores why the current national sample failed to replicate pilot study results. We have divided this discussion into three issues: questions of instruments, questions of differences between national and regional samples, and questions of paradigms as predictors of judging behavior.

The first instance in which one may question the failure of the current studies to replicate previous results pertains to the instruments employed. The primary change made on the questionnaire was to add valence to choices of decision discriminants. In the pilot study, a respondent could indicate his or her strength of belief by reporting the importance of an element in judging. What the respondent could not tell us, however was the direction of the discriminant's influence (e.g., are counter-intuitive arguments helpful or harmful?). The addition of choice of valence (whereby respondents could indicate whether an element "helped" or "hurt") was intended to refine responses. Instead, we may have confused some respondents.

Comments on questionnaires--question marks, etc.--suggested that some subjects did not understand the additional dimension of evaluation for discriminants used for 28 Likert scale items.

Second, the current studies may be inclusive in part because coding categories may need further revision. As we coded ballots, we noticed that we had not devised an exhaustive set of discriminants. We also noted that in some instances the categories we had devised were not mutually exclusive. Coding ambiguity could have minimized the identification of true effects by permitting the miscategorization of discriminants.

Third, we believe that the workload of the content analysis effort contributed to the non-identification of true discriminants. Two hundred and seventeen ballots from twenty-three critics yielded 934 judgments. Similar coding protocols were required for judge philosophy statements. Coding effects (fatigue, drift, etc.) are likely under these circumstances.

Evaluation and revision of instruments is warranted. Categories should be exhaustive and exclusive. Coders need to operate from the same set of assumptions. Inter-coder reliability estimates need to remain realistic. We shall continue to reject "boosting" reliability estimates by refusing to include unused categories in such estimates. We don't believe that mutually non-selected categories should be treated as "inter-coder agreement."

The second set of issues concerns differences obtained in the regional pilot study versus those observed in the national

study. The regional pilot sample yielded more discriminants associating philosophy statements and questionnaires with ballot elements. We fully expected to replicate and expand the description of paradigm taxonomic elements. Instead, we found fewer distinct elements. Part of the boundary definition problem is attributable to the apparent merger of paradigm elements. Tabula rasa merged with three other paradigms on at least six discriminants when measured against seven key discriminators. It also merged with all other paradigms except Stock Issues on five of seven discriminants. Aggregate rankings of paradigms showed that several were clustered.22

Some differences between the regional pilot and national sample may reflect varying assimilation effects that operate at regional versus national tournaments. Regional tournaments are populated largely by critics who interact regularly with each other (directly through conversation and indirectly through ballots written for each other's students). Such interaction may move the debate activity toward an assimilation of standards. But when national samples are analyzed, the same cohesiveness is less likely. First, the national sample may merely aggregate several separate (and different) regional samples. Mixing them together into a common data pool may not result in assimilation. Second, even if there were a "national" standard that judges impose upon themselves (as distinct from the way in which they behave when they are at regional tournaments), the larger distribution of participants in a national sample increases the

likelihood of deviant (non-assimilated) critics appearing in the judging pool.

Finally, we believe our results suggest that judge philosophies do not predict judge behavior because judges do not apply professed beliefs in debate round evaluation. One CEDA judge devoted his philosophy statement to deriding the premise that philosophies either reflect a critic's beliefs or could predict a critic's behavior.23 Several findings in the present study and from the pilot make it plausible to question whether either philosophies or paradigms are applied in any consistent fashion.

First, as unstructured critic assessments of belief, philosophy statements impose the least constraint of any of the instruments. Judges have the latitude to express their preferences in nearly any manner they see fit (including denial of the legitimacy of the philosophy statement).

Second, in both the pilot and present study, respondents' questionnaire preferences were recorded as direct responses. No interpretation of their answers was required. The current study validated the questionnaire as an instrument for obtaining critics' preferences.

With two separate instruments (philosophy statements and questionnaires) recording critics' preferences, it is legitimate to question whether these self-report instruments are reliable indicators of behavior. We believe that judges tend to write philosophy statements that reflect conventions acceptable within

the forensics community. Because of the great variability from round to round, judges are under little scrutiny to implement these conventions in any systematic fashion. Decisions reflect round specific ad hoc impressions that may bear only facial similarity to the larger organizing principles explicit in the judge's philosophy statement, and correspond even less to general paradigm requirements. The present study's failure to identify distinctive paradigm taxonomic elements is evidence for the non-existence (or at least non-distinctiveness) of paradigms. We offer three explanations.

First, while paradigms exist conceptually, they don't necessarily possess distinctive boundaries. Judges employ the label for a paradigm, but aren't obligated to adhere to any standard definition or use convention. So a judge may be "Tabula rasa" (whatever that means) and something else. The high degree of overlap observed for research question #2 in the present study (as well as similar unclear boundaries in the pilot (1989a) evidence fuzzy boundaries). In addition, the overwhelming majority of CEDA judges are willing to employ a paradigm other than that which they prefer if so requested by debaters. It should not be surprising under these circumstances that paradigms operate only as labels delimiting criteria.

A second explanation for the failure of paradigms to predict judges' behavior is that while paradigms exist, they are not distinctive within CEDA. Hence, judges don't know how to apply them. Many traditional paradigms have their origin in policy

debate (Stock Issues, Hypothesis Testing, etc). If NDT debate is to be criticized, it may be criticized for its generation of multiple perspectives (paradigms) by which debate issues may be resolved. CEDA's problem is the opposite. It has no single consensual set of standards by which debates are to be adjudicated. Consequently, the NDT-based models for resolving debates are force-fit upon CEDA rounds (for which they were not intended).

Finally, assuming that paradigms do exist (with distinctive boundaries), one may question whether judges truly understand them. Employing a common paradigm label does not compel the user to pass a qualifying exam in the use of the paradigm. Just as Democrats may reflect a range of political opinions that range from very conservative to very liberal, so it may be that paradigms attract adherents to a common label, but with very different underlying core beliefs.

Regardless of the reasons for paradigm definition failure, the implication is to call into question the method of relying upon self-reports of judging preference as a valid and reliable indicator of subsequent judging behavior. Previous investigations which claim to identify paradigms, philosophies, or patterns of preference should be questioned because of the absence of consistency between "professed belief" statements and actual behavior in the current study.

Continued research investigating the relationship between expressed preferences and subsequent behavior in debate judging

is clearly warranted by this study. If further research fails to establish a consistent relationship between paradigm claimed on judging philosophies and actual ballot behavior, then it may be necessary to re-evaluate the pedagogical benefit of promoting judge philosophy statements.

SYRACUSE DEBATE UNION JUDGING CRITERIA QUESTIONNAIRE

Instructions: Please circle responses.

At left, indicate how much each element should influence decisions; at right, indicate whether presence of the element should help or hurt a team's prospects for winning the round.

[none --> a lot]							
1	2	3	4	5	counter-intuitive arguments	help	hurt
1	2	3	4	5	counter-warrants	help	hurt
1	2	3	4	5	evidence attacks	help	hurt
1	2	3	4	5	evidence out of context	help	hurt
1	2	3	4	5	lack of evidence	help	hurt
1	2	3	4	5	non-applicable evidence	help	hurt
1	2	3	4	5	lack of topicality	help	hurt
1	2	3	4	5	fulfill aff burden of proof	help	hurt
1	2	3	4	5	quality of analysis	help	hurt
1	2	3	4	5	new arguments in rebuttals	help	hurt
1	2	3	4	5	points made during cross-ex	help	hurt
1	2	3	4	5	adherence to time limits	help	hurt
1	2	3	4	5	affirmative fiat of key points	help	hurt
1	2	3	4	5	arguments about debate theory	help	hurt
1	2	3	4	5	repugnant values	help	hurt
1	2	3	4	5	absence of values	help	hurt
1	2	3	4	5	theoretical arguments	help	hurt
1	2	3	4	5	dropped arguments or issues	help	hurt
1	2	3	4	5	justification arguments	help	hurt
1	2	3	4	5	significance arguments	help	hurt
1	2	3	4	5	inherency arguments	help	hurt
1	2	3	4	5	presentation skills	help	hurt

At left, indicate how much each element should influence speaker points; at right, indicate whether presence of the element should help or hurt a debater's rank in the round.

[none --> a lot]							
1	2	3	4	5	speed of presentation	help	hurt
1	2	3	4	5	eye contact with judge	help	hurt
1	2	3	4	5	pacing of presentation	help	hurt
1	2	3	4	5	use of inflection	help	hurt
1	2	3	4	5	obnoxious behavior	help	hurt
1	2	3	4	5	tag team practices	help	hurt

Please rank (1-10) the importance of paradigms you routinely apply in your decisions. (1 = highest rank)

Argument critic	____	Value comparison	____
Argument skills	____	Judicial model	____
Public audience	____	Policy implications	____
Hypothesis testing	____	Stock issues	____
Tabula rosa	____	Other	____

(specify: _____)

Do you ever ask to inspect evidence? Y N

Will you discuss your decision or ballot comments with debaters immediately after a round? Y N

What percent (0-100) of your typical ballot comments and decision criteria are devoted to each of the following? (Each column should sum to 100.)

	Comments	Decision Criteria
Substantive remarks	_____	_____
Procedural remarks	_____	_____
Presentation remarks	_____	_____

Should Affirmative points which are not specifically countered by Negative be held as proven? Y N

What percent of your ballots include low-point wins? ____

On your ballots, what is the typical spread in speaker points between the winning and losing teams? ____

What is the relative importance of these objectives of debate?

[useless --> vital]

1	2	3	4	5	development of speaking skills
1	2	3	4	5	development of logical reasoning
1	2	3	4	5	familiarity with research techniques
1	2	3	4	5	improved organization

How many years have you judged intercollegiate debate?

0-2 3-5 6-8 9-11 12-14 15-17 18-20 20+

What percentage of the rounds you have judged have been NDT?

0-9 10-19 20-29 30-39 40-49 50-59 60-69 70+

How many tournament debate rounds have you judged during the past three semesters?

0-16 17-32 33-48 49-64 65-96 97-128 128+

How many years have you coached intercollegiate debate?

0-2 3-5 6-8 9-11 12-14 15-17 18-20 20+

How many semesters of debating experience have you had personally, in high school and college?

none 1-2 3-4 5-6 7-8 9-10 11-12 13-14 15+

Do you hold a degree in speech, drama, journalism or communications? Y N

Do you hold an appointment as a college faculty member (other than as a graduate assistant)? Y N

Please print your first and last names. (Note: Names will be used for analysis only, not for reporting results.)

first _____ last _____

CODING CATEGORIES FOR BALLOT COMMENTS Acq.# ____

Critic _____ Ballot # _____ Coder _____

I. **MATRIX** - The written portion of the ballot should be categorized in the following matrix as a percentage of the total in 10% increments:

0 = 0 - 9 %	5 = 50 - 59 %
1 = 10 - 19 %	6 = 60 - 69 %
2 = 20 - 29 %	7 = 70 - 79 %
3 = 30 - 39 %	8 = 80 - 89 %
4 = 40 - 49 %	9 = 90 -100 %

A. Criticism Commentary: Presentation Elements _____

B. Criticism Commentary: Substantive Elements _____

C. Decision Criteria: Present in Decision _____

D. Decision Criteria: Rejected in Decision _____

II. **JUDGING PARADIGM** - Code each judging paradigm as

1 = mentioned in decision criteria
0 = not mentioned in judging criteria

E. ____ Tabula Rasa

F. ____ Value Comparison

G. ____ Policy Implications

H. ____ Argument Skills

I. ____ Argument Critic

J. ____ Stock Issues

K. ____ Public Audience

L. ____ Hypothesis Tester

M. ____ Judicial Model

N. ____ Other (_____)

III. **DISCRIMINANTS** - Code the following items on the written portion of the ballot:

0 = not present
1 = present in commentary with positive valence
2 = present in commentary with negative valence
3 = present in decision
4 = rejected in decision

O. ____ Topicality

P. ____ Justification

Q. ____ Significance

R. ____ Inherency/Causality

S. ____ Uniqueness/Intrinsic

T. ____ Issue Default/Dropped

U. ____ Turn-around

V. ____ Cross-Application

W. ____ Case Coverage

X. ____ New Argument

Y. ____ Evidence Source Quality

Z. ____ Cross Examination

AA. ____ Squirrel Case

AB. ____ Generic Argument

AC. ____ Counter-Warrants

AD. ____ Obnoxious Behavior

AE. ____ Quality of Analysis

AF. ____ Burden of Resolution

AG. ____ Prima Facie

AH. ____ Burden of Rejoinder

AI. ____ Burden of Proof

AJ. ____ Common Sense/Counter -Intuitive Arguments

AK. ____ Evidence Context

AL. ____ Evidence Applicable

AM. ____ Evidence Sufficiency

AN. ____ Ethics

AO. ____ Delivery

AP. ____ Organization

AQ. ____ Time Limits

AR. ____ Debate Theory Arg.

JUDGE PHILOSOPHY CODING CATEGORIES Seq.# ____

Critic _____ Coder _____

I. **MATRIX** - The content of the philosophy should be categorized into two dimensions: Philosophy which deals with "Presentational" elements and that which deals with "Substantive" elements. Use the following range increments:

0 = 0 - 9 %	5 = 50 - 59 %
1 = 10 - 19 %	6 = 60 - 69 %
2 = 20 - 29 %	7 = 70 - 79 %
3 = 30 - 39 %	8 = 80 - 89 %
4 = 40 - 49 %	9 = 90 -100 %

A. Presentational Elements _____

B. Substantive Elements _____

II. **JUDGING PARADIGM** - Code each judging paradigm as

0 = not mentioned in philosophy statement
1 = mentioned in philosophy statement

C. ____ Tabula Rasa

D. ____ Value Comparison

E. ____ Policy Implications

F. ____ Argument Skills

G. ____ Argument Critic

H. ____ Stock Issues

I. ____ Public Address

J. ____ Hypothesis Tester

K. ____ Judicial Model

L. ____ Other (_____)

III. **DISCRIMINANTS** - Code the following items from the Philosophy

0 = not mentioned in philosophy statement
1 = mentioned in a positive valence (i.e., "like," "good," etc.)
2 = mentioned in a negative valence (i.e., "dislike," "bad," etc.)

O. ____ Topicality

P. ____ Justification

Q. ____ Significance

R. ____ Inherency/Causality

S. ____ Uniqueness/Intrinsic

T. ____ Issue Default/Dropped

U. ____ Turn-Around

V. ____ Cross-Application

W. ____ Case Coverage

X. ____ New Argument

Y. ____ Evidence Source Quality

Z. ____ Cross Examination

AA. ____ Squirrel Case

AB. ____ Generic Arguments

AC. ____ Counter-Warrants

AD. ____ Obnoxious Behavior

AE. ____ Quality of Analysis

AF. ____ Burden of Resolution

AG. ____ Prima Facie

AH. ____ Burden of Rejoinder

AI. ____ Burden of Proof

AJ. ____ Common Sense/Counter Intuitive Arguments

AK. ____ Evidence Context

AL. ____ Evidence Applicable

AM. ____ Evidence Sufficiency

AN. ____ Ethics

AO. ____ Delivery

AP. ____ Organization

AQ. ____ Time Limits

AR. ____ Debate Theory Args.

REFERENCES

Brey, J. (1989). "A Descriptive Analysis of CEDA Judging Philosophies Part I: Definitive Acceptance or Rejection of Certain Tactics and Arguments." CEDA YEARBOOK 10: 67-77.

Bryant, M. (1983). "A Content Analysis of Selected NDT and CEDA Transcripts." Unpublished. Paper presented at the Third SCA/AFA Summer Conference on Argumentation. Alta, Utah.

Buckley, D. (1983). "A Comparison of Judging Paradigms." ARGUMENT IN TRANSITION: PROCEEDINGS OF THE THIRD SUMMER CONFERENCE ON ARGUMENTATION. Eds. Zarefsky, D., Sillars, M., and Rhodes, J. Annandale, VA: Speech Communication Association, 858-870.

Cox, J. (1974). "A Study of Judging Philosophies of the Participants of the National Debate Tournament." JOURNAL OF THE AMERICAN FORENSIC ASSOCIATION (JAFA) 11: 61-71.

Cross, J., and R. Matlon (1978). "An Analysis of Judging Philosophies in Academic Debate." JAFA 15: 110-123.

Dudczak, C., and D. Day (1989a) "A Profile of CEDA Debate Critics," SPHERES OF ARGUMENT: PROCEEDINGS OF THE SIXTH SCA/AFA SUMMER CONFERENCE ON ARGUMENTATION. Ed. Gronbeck, B. Annandale, VA: Speech Communication Association, 427-433.

Dudczak, C., and D. Day (1989b) "A Taxonomy of CEDA Debate Critics." Unpublished. Paper presented to the Speech Communication Association Conference, San Francisco. (ERIC # ED 318 071)

Gaske, P., D. Kugler, and J. Theobald (1985). "Judging Attitudes and Paradigmatic Preferences in CEDA Debate: A Cumulative and Construct Validity Investigation." CEDA YEARBOOK 6: 57-66.

Henderson, B., and D. Boman (1983). "A Study to Determine if Debate Judges' Philosophy Statements are Consistent with Their Later Related Ballot Statements." JAFA 19: 191-198.

Hollihan, T., P. Riley, and C. Austin (1983). "A Content Analysis of Selected CEDA and NDT Judges' Ballots." ARGUMENT IN TRANSITION Eds. Zarefsky, D., Sillars, M., and Rhodes, J. Annandale, VA: Speech Communication Association, 871-882.

Holsti, O. (1969). CONTENT ANALYSIS FOR THE SOCIAL SCIENCES AND HUMANITIES. Reading, MA: Addison-Wesley

Klecka, W.R. (1980). DISCRIMINANT ANALYSIS. Beverly Hills, CA: Sage Publications.

Krippendorff, K. (1980). CONTENT ANALYSIS: AN INTRODUCTION TO ITS METHODOLOGY. Beverly Hills: Sage Publications.

Lee, R., K. Lee, and M. Seeger (1983). "Comparison of CEDA and NDT: Differences in Program Characteristics and Forensic Director Attitudes." ARGUMENT IN TRANSITION Eds. Zarefsky, D., Sillars, M., and Rhodes, J. Annandale, VA: Speech Communication Association, 845-57

Paisley, W. (1969). "Introduction," THE ANALYSIS OF COMMUNICATION CONTENT. Gerbner, G. et al, eds. New York: Wiley, 286.

Thomas, D. (1977) "Analysis of Judges and Debaters at NDT" Unpublished. Data collected from the 1977 National Debate Tournament. Used with permission.

Webb, E. and K. Roberts. (1969) "Unconventional Uses of Content Analysis in Social Science," THE ANALYSIS OF COMMUNICATION CONTENT. Gerbner, G. et al, eds. New York: Wiley, 319-332.

ENDNOTES

1. Brey identified the percentage of critics categorized by paradigm preference and then separately reported elements of judge preference (i.e., Prefer vs. Abhor "spread"). One cannot determine from his data whether these judge preferences divide along paradigm boundaries.

2. His results are contaminated by a failure to control for differences in time format (i.e., NDT used 10-5 while CEDA used 8-4) and competitors' skill levels (i.e., at NDT finals vs. at CEDA regional tournament).

3. Three elements confound Hollihan et al's findings. First, they treated NDT and CEDA judges as aggregate types. NDT judges were categorically compared with CEDA judges without evaluating whether there were within group differences. It is questionable whether this assumption is true given the previous research establishing "paradigm" types within each respective debate format. Second, at the time of Hollihan et al's research CEDA had not instituted its National tournament (with its accompanying judge philosophy requirement). The absence of a critic philosophy requirement in CEDA would tend to reflect itself in less well-formulated judging standards. Third, since NDT debaters had access to judge philosophy statements, they theoretically should have been better able to adapt to their critics' preferences, minimizing commentary generated by their critics. CEDA debaters, less informed of their critics' preferences, would theoretically be less adaptive to their critics' expectations. This in turn would create a relatively greater need for critics to provide commentary retroactively to explain their judging preferences.

4. Henderson and Boman failed to conform to several validity and reliability standards. Primary is their violation of exhaustiveness in content analysis. Only items which appeared on both the judge philosophy statement and the ballot were coded for consistency. One cannot determine whether some professed preferences were inconsistent because the critic choose not to articulate them on the ballot. For instance, a critic who professed to vote on inherency could only be coded as inconsistent if s/he expressly contradicted the philosophy statement by writing on the ballot something to the effect that "I don't vote on inherency." The failure to address inherency on the ballot would not have been coded, but any recognition of inherency in the decision would have been coded as consistent. Other problems surround the use of a single ballot for 19 of 23 usable critics.

5. Dudczak and Day generated a consistency index by comparing (1) critics' professed preferences measured through two instruments (judge philosophy statements and a survey questionnaire) with ballot comments. Since a critic would need to demonstrate consistency across three items (instead of the two used by Henderson and Boman), some lower consistencies reported by Dudczak and Day may be an artifact of differences in analytic procedure.

6. These findings are of limited utility since the pilot employed subcritical numbers of subjects. However, unlike the Henderson and Boman analysis (which largely relied upon the analysis of a single ballot from each subject), Dudczak and Day used multiple ballots per subject (the average was 13.1 ballots/subject, with a threshold minimum of 6 ballots).

7. The generalizability (national vs. regional) and sample size of the present research are expected to influence these and other results of the pilot.

8. While the hypotheses are stated here in the direction of anticipated results, they were tested as null hypotheses.

9. Paradigms were merged into meta-paradigm groups in the pilot study because of the limited number of subjects representing each paradigm. The hallmark of "audience-centered" paradigms is the expectation that speakers would adapt their presentation content and style to audience preferences.

10. More philosophies were available than were used in the study. However, since we were interested in comparing professed philosophies with other professions of belief (as indicated on questionnaires) and with actual behavior (as shown on ballots), only subjects for whom we had all three types of documents were used in this part of the study. We assumed that judging philosophies are relatively stable. Hence, while we invariably used the most recent philosophies available, we also employed philosophy statements taken from earlier tournament books when no more recent statements were available. The oldest statement came from the 1987 National tournament booklet.

11. The valence choices for Likert Scale items allowed evaluation of both the strength of belief and the polarity of the belief.

12. All twenty-nine tournament directors solicited agreed to administer and return the questionnaires. Only two of the eighteen who did not follow through offered explanations for their non-return (both involving the ostensible efforts of over-zealous janitors). The non-returns created a

substantial problem since many critics, having completed a survey at an earlier tournament, were unwilling to complete another survey at a subsequent tournament. The direct mailing solicitation yielded a 48% response (17 of 35), although one of the questionnaires was received too late to be included in the current analysis.

13. The 137 unusable ballots included 68 blank ballots, 13 illegible ballots, 21 round forfeits, 22 judge disqualified (i.e., a member of the research team), 6 "oral critiques", 5 "useless comments", and 2 duplicate ballots.

14. Only a single coder's results are reported in this manuscript. The study protocol specifies that a second coder is scheduled to code ballot comments independently to establish appropriate inter-rater reliability estimates.

15. The method used to calculate the correlation coefficient was to sum the product of inter-coder correlations, multiply the result times the number of times the category was employed, then divide the products by the total number of coding judgments made. This technique provided a weighted model representing the agreement times frequency of category use. We believe the integrity of this method diminishes inflated reliability calculations created when coders treat mutual non-selection of a category as "agreement."

    It should also be noted that with 934 separate comparisons made by two coders on 23 philosophies (about 20 per coder per philosophy), the treatment of the non-selected categories as "agreement" would have inflated the reliability coefficient to at least (R = .75).

16. "Justification" and "organization" discriminants transcend paradigms, unless the two paradigms with similar correlations in each case are overlapping (Policy Implication and Argument Critic; Argument Critic and Stock Issues). However, it still is possible that these paradigms are distinct, merely sharing two relatively strong discriminatory components.

17. One cannot rule out the possibility that one or more of the paradigm pairs masks differences among paradigms, creating the impression of a false commonality. For instance, tabula rasa combines with three other paradigms. If the threshold for atypical similarity were five of seven discriminators, tabula rasa would combine with each other paradigm except "Stock Issues." This may well suggest that tabula rasa operates as a "meta-paradigm."

18. The pilot study had found support for this hypothesis

19. The pilot study also showed support for this hypothesis.

20. Analytic-centered paradigms included Stock Issues, Value Comparison, Hypothesis Testing, Policy Implications, and Judicial Model. Audience-centered paradigms included Public Audience, Argument Skills, and Argument Critic. Creation of these two "meta-paradigms" placed emphasis on resolving issues analytically vs. in presentational terms.

21. While not tested as a hypothesis, the greatest association with low-point wins was years of experience coaching debate ($r = .137$). This finding also was not significant, however.

22. The univariate mean ranks for paradigms ranked on the questionnaire were:

Argument Critic	3.26
Tabula Rasa	3.29
Value Comparison	3.56
Argument Skills	4.12
Stock Issues	4.95
Policy Implications	5.24
Hypothesis Testing	5.73
Judicial Model	6.33
Public Audience	6.92

23. See Todd Graham, 1990 CEDA Judge Booklet